ARCHITECTURE AND DESIGN PROCESS OF THE INDIVIDUALIZED ASSESMENT SYSTEM INTEGRABLE TO DISTANCE EDUCATION SOFTWARES

Hacer OZYURT
Ozcan OZYURT
Prof. Dr. Adnan BAKI
Karadeniz Technical University
Fatih Faculty of Education,
Trabzon, TURKEY

ABSTRACT

Assessment is one of the methods used for evaluation of the learning outputs. Nowadays, use of adaptive assessment systems estimating ability level and abilities of the students is becoming widespread instead of traditional assessment systems. Adaptive assessment system evaluates students not only according to their marks that they take in test exams but also according to their ability levels.

In this study, we dealt with theoretical background of the adaptive assessment systems. Study covers the structure, characteristics and architecture of the computerized adaptive test systems in detail. It dwells on the Item Response theory which is used for improving computerized adaptive test systems and the models with which this theory is used. Besides, processing steps which are required to realize a computerized adaptive test system and algorithms used for the validation of the test systems are introduced in detail.

Briefly, study introduces the structure and the improvement processes of the computerized adaptive testing systems which can be integrated to distance education software.

It is possible to say that integration of these systems into distant education software allows individualized learning and individualized assessment studies to be carried out successively in distant education software as well.

Keywords: Individualized Adaptive Assessment, Assessment Modules in Distance Education, Adaptive Testing System, Individualized Learning.

INTRODUCTION

Assessment is an inseparable part of the learning process. we can get information both about the learning process and ability levels of students via assessment (Okonkwo, 2010). And by this means, assessment can contribute to improvement of the education quality using the feedback of the students (Peterson & Irving, 2008; Rastgoo & Namvar 2010). In recent years, researches made in the fields of individualized learning and individualized assessment lead to the creation and development of the adaptive assessment systems.

The key word of the adaptive assessment is the assessment of the ability level and the learning process of the students in a more effective and better way (Gouli, Kornilakis, Papanikolaou & Grigoriadou, 2001; Sitthisak, Gilbert & Davis, 2007). Adaptive assessment systems are presenting an individualized assessment environment to the students. Thanks to this, it becomes possible for the students to see their own skills and learning outputs better.

In adaptive assessment systems, selection of the following question is changing dynamically according to the performance of the student. And by this means each student is being assessed with a test specific for itself according to its skill level. Thanks to Adaptive assessment students are being tested according to their competence. And this situation will prevent the students from being boried while they are answering the questions.

Hence it is taken under guarantee that with the adaptive assessment system the students will not encounter with neither very difficult nor very easy questions (Lazarinis, Green & Pearson, 2010; Tian, Miao, Zhu & Gong, 2007).

Adaptive assessment systems are generally known as Computerized Adaptive Testing (CAT). In contrast to the classical testing system which gives the same constant test to each examinee, cat systems provide each examinee with a unique test designed for its own skill level. Here the selection of the following question is arranged according to the skill level of the examinee and after the answer of the each question skill level of the examinee is re-estimated.

The selection of the following question is adapted according to this estimated ability level. The computer re-estimates the ability level of the examinee until it reaches the statistically acceptable level or it reaches the maximum question number (Triantafillou, Georgiadou & Economides, 2007).

At the end of the test the final skill level of each examinee is estimated. From the examinees answering the questions in the same number test points of the ones who has high skill level will be higher than the ones whose skill level are low. According to the results of the various researches, the adaptive assessment is generally more effective and efficient than the non-adaptive systems (Antal & Koncz, 2011; Gouli, Papanikolau & Grigoriadou, 2006).

This study deals with the architecture and design of the individualized adaptive assessment module which can be integrated into distant education software. In the following part of the study it is give place to the architecture of this module, algorithm used in module and process about the improvement of the module.

In short proceedings process of a CAT application is introduced from the very basic to realization in detail.

THEORETICAL BACKGROUND OF COMPUTERIZED ADAPTIVE TESTING (CAT) SYSTEMS

The first example of CAT practice in computer environment is seen in Alfred-Binet IQ test. This test was developed for children in 1905.

It begins with a beginning question according to the age of each child, difficulty degrees of the following questions are determined according to the answer and the test is finished in the event that the test is failed for a few times (Georgiadou, Triantafillou & Economides, 2006; Fetzer et. al., 2008). Although this test is seen very simple when it is compared to other adaptive tests in our day, it constitutes the basic of these practices

Despite the fact that the generation of the adaptive test in computer environment is at the very beginning of the 1900s, in the following years there were no interests for this test and its using did not become widespread. Defense ministry of USA made studies on the CAT to use it in land, air and naval forces. But because of the cost, size and speeds of the computers, efficient benefit was not received. In 1960s and 1970s, a rapid development was seen in the adaptive test practices. This development was realized depending on the increasing studies in the field of item respond theory in addition to the new improvements in computer technology (Weiss, 2004; Fetzer et al., 2008). In 1980s more developed methods began to be applied for the CAT practices. For example, variable step- dimensional item selection models benefitted from the technical matters such as providing item selection methods at indefinite number and estimation of interactive general information of the examinees.

CAT Systems and Their Features

Cat systems have four main components. These are item pool, item selection procedure, information estimation, and termination rule. The key concept of the CAT is the best match of a test item to the ability level of the examinee (θ). Item difficulty and the most appropriate equivalence provide maximum information about the ability level of the examinee.

The success of CAT substantially depends on creating a good item pool. Two criteria are searched in test items in order to be appropriate for the CAT functions. The first one is to create a huge item database composed of items whose item characteristics are known, the other one is that all the items should be uniplanar (Boyd, 2003; Fetzer et al., 2008). The final item pool should be composed of 5 to 10 times more items than the items to be present to the examinee. According to this it is advised that for a thirty item test there should be 150-300 items in the item pool (Georgiadou, Triantafillou & Economides, 2006; Fetzer et al., 2008). After the beginning item pool is created, data for each item are collected. Even though for the typical Item Response Theory (IRT) analysis at least 300 samples are required, more than 500 samples are preferred. Items should be separated to smaller subordinate pool and accurate data should be collected.

Item parameters;

- **discrimination**
- **difficulty**
- **guessing can be estimated.**

These items are used for deciding which items will be put to final pool. Final item pool is integrated to CAT system and an item selection way appropriate for examines is created (Weiss, 2004; Fetzer et al., 2008).

CAT systems provide each examinee with a test that is adapted to its own ability level. Besides giving fixed length test to each examinee makes item selection adaptation appropriate to ability level alternately. It is because asking questions to a person above of his/her ability level (very difficult) or below from his/her ability level (very easy) does not give information about the ability level of the examinee. Moreover asking questions which are not appropriate for his/her ability level also bores the examinee (Georgiadou, Triantafillou & Economides, 2006; Tian et al., 2007; Fetzer et al., 2008). In this context, It is required that there should be efficient questions for each ability level in question pool to be created. This also changes according to the parameters of the questions (difficulty level, distinctiveness index and guessing factor etc) in the pool. It is also an important matter to determine the parameters of the questions in the pool. Due to the fact that the questions are applied to hundreds even thousand people before they are added to the pool, the parameters are known and the questions to be applied are selected with regard to these parameters (Boyd, 2003).

In CAT systems, after each answer information estimation is updated and the following item is selected according to the new estimation and also appropriate characteristics. First of all, in order to determine the beginning level of each person an item in average difficulty is presented in CAT. During the testing process each answer is estimated quickly, if examinee answers the question correctly, testing system statistically determine the examinee as high level. Similarly, if examinee answers the question incorrectly, the system determines it as low level. Testing system selects the most appropriate question for its competence level estimated in that moment and presents to the examinee. If it answers the following question correctly, its ability level is estimated as high but if it answers the question incorrectly its ability level is estimated as low. Testing system presents the following item in response to new information estimation. This circle continues until the information estimation of the examinee reaches estimation in statistically acceptable certainty or it reaches a limit such as maximum testing item (Tian et al., 2008).

All the CAT systems should combine some types of termination rules in order to stop the item selection/information estimation circle. In this situation the rules can be composed of one or several of the rules below (Boyd, 2003; Weiss, 2004; Fetzer et al., 2008).

- **Giving minimum item number.**
  - **This rule is in the first priority, because none of exams can be finished prior to reaching minimum item number.**
- **Giving the maximum item number**
- **Both the item and the reaching of testing level to the maximum time limit.**
- **Reaching acceptable certainty level for the ability estimation level.**
- **When ability estimation moves away efficiently from the pass-fail criteria.**
- **When examinee behaves out of the test.**
  - **CAT program is able to understand if the answer ranks are very slow or very quick. It applies to the testing manager for the final decision whether postponing the test or stopping it.**

In CAT point is determined according to the difficulty level of the questions that the examinee answers correctly.

As a consequence, even though all the examinees answer the same percent question, examinees in high ability level will get better points because they answer more difficult questions correctly (Tian et al., 2008).

Item Response Theory (IRT)

IRT is a mathematical model that takes into consideration the possibility for the examinee's giving the correct answer to each item and defines the examinee independently from examinee and the test. In this theory, even if two different tests including different questions are applied to the same person, the estimated ability level is not different in both tests. The point which is reached at the end of the test is not the test point but an ability level estimation of the examinee. This ability level is known as theta (θ) and it gets a value between the -3 and +3. On the (θ) scale, zero shows the average ability level, negative values show ability level lower than the average, and positive values show ability level higher than the average (Boyd, 2003; Weiss, 2004). As in every mathematical model IRT has also some certain acceptances (Fetzer et al., 2008). These acceptances are unidimensionality and local independence.

Unidimensionality

For an IRT model aiming to measure ability level of the examinee appropriately test item intended only for a unique structure or implicit features should be arranged. That is, In the Unidimensional IRT models it is accepted that test performance is determined by a unique latent variable (information). It is impossible to fulfill this assumption completely. Because it is known that test anxiety, testing strategies, personality, ability to make operation quickly, motivation and several other similar variables effect the test performance. But it is expected for these other variables to be the lowest level. It is basic for this assumption that seeing that a dominant factor or component affects the test performance and explains it on a large scale. This dominant component is wanted to be the variable measured by the test. Factor analysis methods can be used for testing unidimensionality. The portion of first factor variance to the second factor variance is advised as a unique unidimensionality index. If there is a marked difference between the first factor and the second factor, this means that unidimensionality is achieved. It is put forward as a method in item selection to apply factor analysis again to the items that is contrary to the factor analysis after they are taken out from the item pool and to eliminate the items in that way until the intended unidimensional structure is reached.

Local independence

Local independence is that the answers that the measuring machine gives to its items are statistically independent when the ability level of the persons is accepted stable (θ). In IRT models it is assumed that the answer that a person gives to different items of the test is statistically independent. Besides, Achieving the Unidimensionality also means achieving the Local independence. In IRT, distribution function of the possibility of giving correct answer to a question according to the ability level is named as Item Characteristic Curve (ICC) and different mathematical equations defining this curve composes the models (Guzmán, Conejo & García-Hervás, 2005).

The models that allows to mark the answer that the persons gives to the test items as true (1)- fal

The most used models used in parameter (Boyd, 2003; Fetzer et al., 2008).

Single parameter logistic Model (1 PL Model)

Known as the Rasch model, this model estimates the skill of the examinee as (θ) one parameter. This parameter is the item difficulty expressed with b. Single parameter model creates a relationship between the item difficulty parameter b and ability level of the person.

This model has two additional assumption such as that the luck success is zero and each item has equal selectivity power. This model uses a logical function containing both item difficulty level and the information estimation of the examinee θ regarding to getting the possibility of the examinee's answering the item correctly for each item. The real value of the b shows the ability level of the examinee. At this point there is the possibility of examinee's answering the question correctly with %50 luck. Formula of the relation is seen below;

$$P(\theta) = \frac{e^{D(\theta-b)}}{1+e^{D(\theta-}} \quad \textbf{(Eq. 1)}$$

b: difficulty degree of the item
θ: ability level of the person
e: 2,718 (fixed)
D: 1,7 (scaling multiplier)
P (θ): the possibility of the person's with the θ ability level answering the item in the b difficulty degree.

In Figure: 1 the graph of this Eq.1.

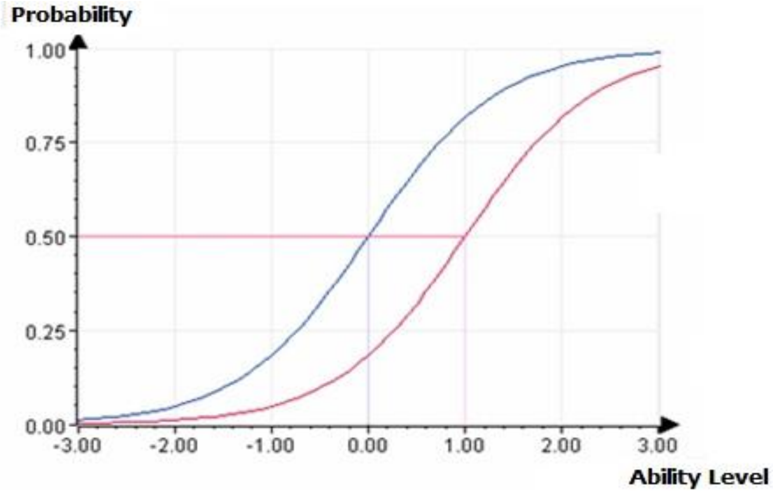

Figure: 1
The possibility of the item's being answered correctly for the ability level estimated according to the item difficulty in one parameter logistic model.

As seen Figure: 1, on one dimension there is item difficulty b and information θ measurement, on the other dimension there is the possibility of item's being

answered correctly. When the item characteristic curve slides to right, the function shows more difficult items.

Two Parameter Logistic Model (2PL Model)

This model takes into consideration a second parameter while estimating the skill of the examinee. This parameter, item distinctiveness or curve slope are showed with a.

This parameter stands for the power of distinction of the low and high information groups of the question in a specific ability level. Item distinctiveness index a is the slope of the item characteristic curve on b point, that is slope of the curve on inflection point. Items whose slope is bigger are more distinctive and more useful.

Formula relation in two parameter model is showed below.

$$P(\theta) = \frac{e^{Da(\theta-b)}}{1+e^{Da(\theta-b)}} \qquad \textbf{(Eq. 2)}$$

a: item discrimination index
b: difficulty degree of the item
θ: ability level of the person
D: 1,7 (scaling multiplier)
e: 2,718 (fixed)
P (θ): The possibility of the person's in the θ ability level answering the item in b difficulty degree correctly.

In Figure: 2 the graph of this relation is showed.

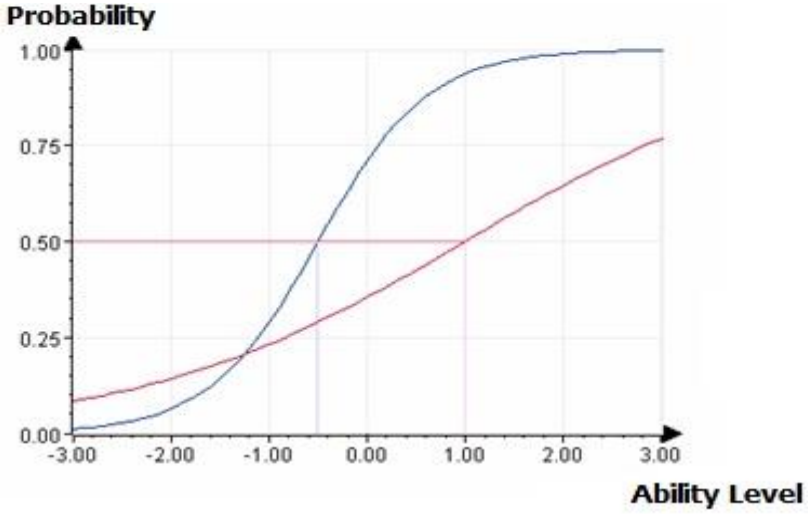

Figure: 2
The possibility of the question's being answered correctly for the information estimation according to the item difficulty and distinctiveness index in two parameter logistic model.

Three parameter Logistic Model (3PL model)

This model includes a third parameter apart from the item difficulty and distinctiveness. This parameter is generally named as guessing parameter and showed with c.

Guessing parameter includes the opinion that relating with the multiple choice tests, the examinee with very low ability level may find the true answer fortunately. The Formula of the relation in three parameter Logistic model can be seen below.

$$P(\theta) = c + (1 - c)\frac{e^{Da(\theta - b)}}{1 + e^{Da(\theta - b)}} \qquad \textbf{(Eq. 3)}$$

a: item distinctiveness index
b: difficulty degree of the item
c: guessing factor
θ : ability level of the person
D: 1,7 (scaling multiplier)
e: 2,718 (fixed)
P (θ): The possibility of the person's in the θ ability level answering the item in b difficulty degree correctly.

Although c value theorically is in the [0, 1] interval, in practice it is in the interval of 0<c<0.35 values. The effect of the three parameters on the information and true answer is seen on Figure 3.

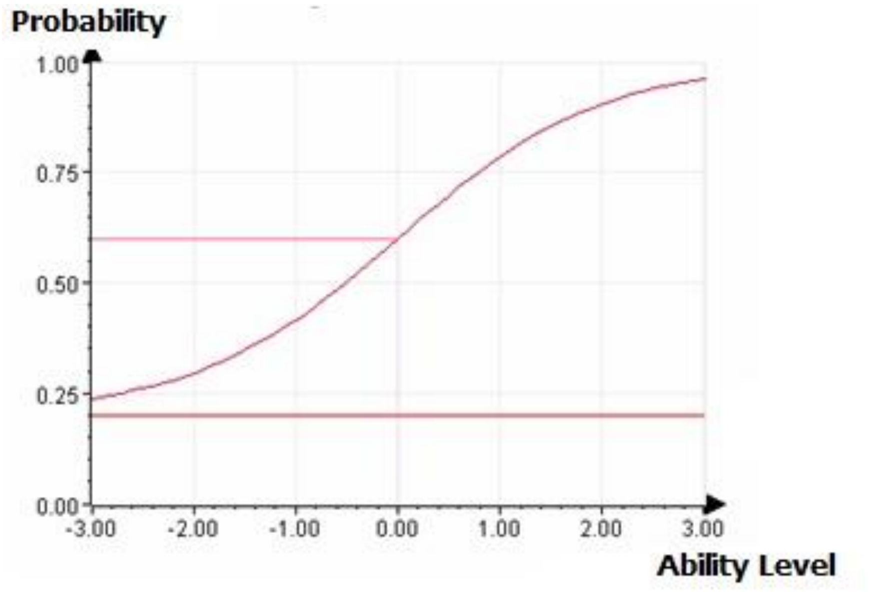

Figure: 3
In three parameters model, the possibility of the item's being answered correctly for the information estimation calculated according to the item difficulty, distinctiveness index and guessing factor.

Calculation of the parameters is difficult because of the complexity in defining the item estimation function for IRT.

For this reason especially when the number of the items and the persons is more, a computer program is needed. There is several software (BILOG, MULTILOG, PARSCALE, RUMM etc.) prepared for this purpose (Raîche, Blais & Magis, 2007). The adaptivity of model and data can be tested by using one of these programs. With this test it can be learned that the data obtained from the test is appropriate for which logistic model. Thus, it can be decided that which logistic model is used in CAT practices.

The Improvement Process and Architecture of a CAT Practice

It is required to follow and fulfill the operation steps below in order to develop a testing system for CAT practice.

- **To create the beginning question pool in numbers between the five and ten times of the test item number.**
- **Applying this question pool 500 and more people and obtaining the test result.**
- **Testing the obtained test results with the model-data harmony and deciding**
- **which logistic model will be used.**
- **Obtaining the parameters (a,b,c) according to the logistic model used.**
- **Obtaining the ICC of each question item appropriate for logistic model used.**
- **Taking out the items that are not appropriate for the logistic model that is used from the item pool.**
- **Creating the final item pool.**

Fulfilling these steps means that the background of the CAT practice is established. The following step is creating the interface of the test and application of the test.

Adaptive Assessment Algorithm for CAT Practice

The adaptive assessment algorithm to be applied in CAT practice can be defined as below:

- **Determine a beginning skill level for the examinees.**
- **Select the most appropriate question for the current skill level estimated for the examinee.**
- **Estimate the skill level according to the answer of the examinee.**
- **Return back to b item and go on until the termination rule is achieved.**
- **Estimate the final skill level of the examinee and calculate the test point according to this skill level.**

According to the description of the algorithm above, four important points come to the forefront:

- **How the most appropriate question for the examinee is selected?**
- **How the skill estimation for the examinee is made?**
- **Which termination rule is used?**
- **How the test point of the examinee is calculated?**

The operation sequences made for these four points are explained in detail in the following paragraph.

Maximum Information Selection (MIS) method can be used for question selection appropriate for the estimated skill level from the question bank. In this method the questions which provide the maximum information about the ability level of the examinee.

The beginning question is selected generally by accepting the ability level of the examinee as midlevel (Boyd, 2003).

The first question is asked from zero skill level. System will give the question which provides maximum information about the skill level in that moment between the skill levels of -3 and +3 to the student according to that whether the answer given by the student is true or false. Thus the student receives the question which is most appropriate for its ability level.

Maximum Likelihood Estimation (MLE) method is one of the most frequently used methods for the ability level estimation of the examinee in CAT practices.

It finds the value of ability level which makes the possibility of answer series as the biggest possibility by relying on fundamentally the answers of the examinee. It is one of the methods whose using question at stable number is widespread in improved tests.

At the end of the tests, the system estimates the answers of the students and test point is calculated according to the IRT.

Test point of the examinee is directly relevant to the estimated skill level. In IRT, the test point is calculated as below after the ability level of the student is determined (Baker, 2001).

(Eq. 4)

TS_j: the test score composed of questions at j number
Θ_j: ability level estimated in j items
$P_i(\theta_j)$: the possibility of correctly answered i: item's answering correctly in estimated skill level.

Figure: 4 shows the operation steps of receiving a test and scoring as a CAT practice are given.

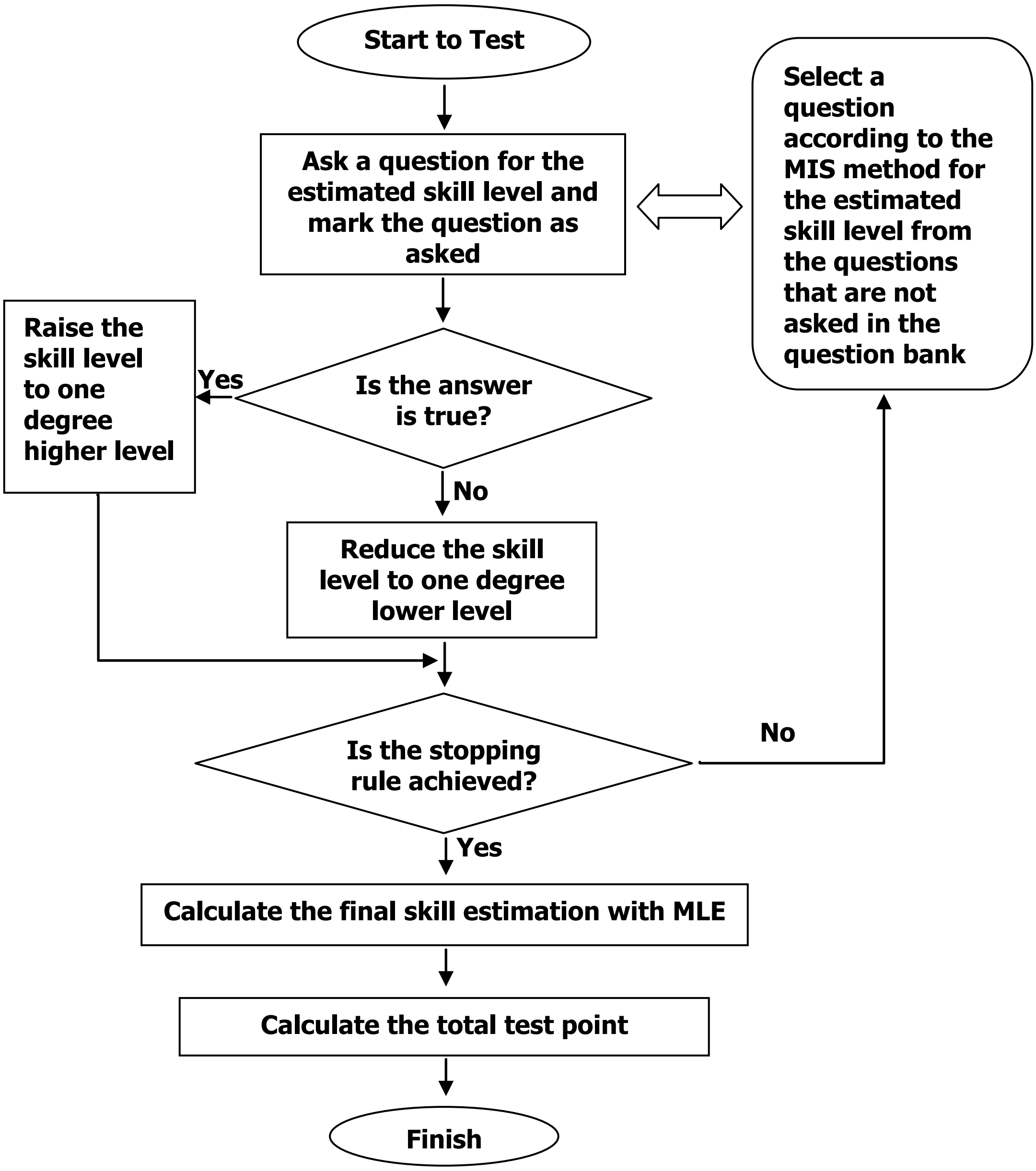
Start to Test
Ask a question for the estimated skill level and mark the question as asked
Select a question according to the MIS method for the estimated skill level from the questions that are not asked in the question bank
Raise the skill level to one degree higher level
Yes
Is the answer is true?
No
Reduce the skill level to one degree lower level
Is the stopping rule achieved?
No
Yes
Calculate the final skill estimation with MLE
Calculate the total test point
Finish

Figure: 4
The architecture of the question selection and ability estimation made in test for CAT practice

CONCLUSIONS and FUTURE WORKS

Measurement is an element of education from the point of evaluating the learning outputs. In parallel with improvements in the fields of computer and software technologies, Computerized Adaptive Testing (CAT) practices have begun to become widespread in the environment of computer-web based learning.

The characteristic of the CAT systems is to present a test to each examinee that is adapted to his/her own ability level. With this structure CAT systems differ from the classical testing practices which give the same fixed test to each student.

Positive results have been obtained when the studies about the using of the CAT systems in education are analyzed. There is positive indications such as that Cat systems and adaptive testing software estimates the skill level of the students trustingly by using fewer question in the measurement/estimation parts of distance education, web- based education and computer aided education in literature (López-Cuadrado et al, 2002; Guzmán, Conejo & García-Hervás, 2005).

The purpose of this study is to introduce the structure and features of the adaptive assessment systems which can be also integrated to the distance education software. Theoretical background, features and improvement process have been introduced.CAT systems can find important application fields to itself especially in distance education systems and in particular web based learning environment. In this context, important benefits can be provided to the field with the improvement and application of the Cat systems.

BIODATA and CONTACT ADDRESSES of AUTHORS

Hacer OZYURT was born in Trabzon, Turkey in 1982. She received the B.Sc. degrees in department of Computer and Instructional Technologies Karadeniz Technical University (KTU) in 2007. Currently, she is a PhD student at Educational Sciences Institute of the Karadeniz Technical University (KTU). Her major research interests are in artificial intelligence in education, adaptive and intelligent tutoring system, computerized adaptive testing and e-learning.

Hacer OZYURT
Karadeniz Technical University, Fatih Faculty of Education,
Trabzon, Söğütlü, 61335, TURKIYE
Email: hacerozyurt@ktu.edu.tr

Ozcan OZYURT was born in Trabzon, Turkey in 1978. He received the B.Sc. and M.Sc. degrees in Computer Engineering from Karadeniz Technical University (KTU) in 1996 and 2000, respectively. Now, He is PhD student at Educational Sciences Institute of the Karadeniz Technical University (KTU). Currently, He is a lecturer in the Department of Computer Technologies at Besikduzu Vocational School at KTU. His major research interests are in the use of artificial intelligence in education, adaptive and intelligent tutoring system, e-learning and mathematics education.

Lecturer Ozcan OZYURT
Karadeniz Technical University, Besikduzu Vocational,
Department of Computer Technologies,
Besikduzu, Trabzon, 61800, TURKIYE
Phone: +90 462 8716922-8562
Email(s): oozyurt@ktu.edu.tr or oozyurt61@gmail.com

Adnan BAKI was born in Trabzon, Turkey in 1960. He completed his bachelor‟s degree in mathematics education at Fatih Faculty of Education, Karadeniz Technical University. He completed his master on curriculum and instruction at University of New Brunswick in Canada, 1990. He completed his doctoral studies (Ph.D.) on teacher education and computer based mathematics teaching at University of London, U.K, 1994. The title of his doctoral research is ―Breaking with tradition in mathematics education: Experiences of Student Teachers within a Computer-based Environment. He has been working as a full-time faculty member, Prof. Dr., in the Department of Secondary Science and Mathematics Education, Fatih Faculty of Education, Karadeniz Technical University, Trabzon, Turkey, since 2004. His research interests are mathematics education, instructional technology, teacher education and distance learning.

Prof. Dr. Adnan BAKI
Karadeniz Technical University, Fatih Faculty of Education,
Department of Secondary Science and Mathematics Education
Trabzon, Söğütlü, 61335, TURKIYE
Phone: +90 462 3777188
Email: abaki@ktu.edu.tr

REFERENCES

Antal, M., & Koncz, S.(2011). Student modeling for a web-based self-assessment system. *Expert Systems with Applications*, 38(6), 6492-6497.

Baker, F. B. (2001). The Basics of Item Response Theory, ERIC Clearinghouse on Assessment and Evaluation, Second Edition.

Boyd, A.M. (2003). Strategies for Controlling Testlet Exposure Rates in Computerized Adaptive Testing Systems, Unpublished Doctoral Thesis, The University of Texas at Austin.

Fetzer, M., Dainis, A., Lambert, S., Meade, A. (2008). Computer Adaptive Testing (CAT) in an Employment Context, PreVisor's PreView.

Georgiadou, E., Triantafillou, E., & Economides, A.A.(2006). Evaluation Parameters for Computer-Adaptive Testing. *British Journal of Educational Technology*, 37(2), 261-278.

Gouli, E., Papanikolaou, K., & Grigoriadou, M.(2006). Personalizing Assessment in Adaptive Educational Hypermedia Systems. *Lecture Notes in Computer Science*, 2347/2006, 153-163.

Guzmán, E., Conejo, R., & García-Hervás, E.(2005). An Authoring Environment for Adaptive Testing. *Educational Technology & Society*, 8 (3), 66-76.

Lazarinis, F., Green, S., & Pearson, E.(2010). Creating personalized assessments based on learner knowledge and objectives in a hypermedia Web testing application. *Computers & Education*, 55(2010), 1732–1743.

López-Cuadrado, J., Pérez, T.A., Vadillo, J.A. & Arruabarrena, R.(2002). *Integrating Adaptive Testing in an Educational System*. First International Conference on Educational Technology in Cultural Context: ETCC2002. Joensuu, Finland: University of Joensuu, 133-149.

Okonkwo, C. A. (2010). Sustainable Assessment and Evaluation Strategies for Open and Distance Learning. *Turkish Online Journal of Distance Education-TOJDE*, 11(4), 121-129.

Peterson, E. R., & Irving, S. E.(2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction*, 18(3), 238-250.

Raiche, G., Blais, J.-G., & Magis, D. (2007). Adaptive estimators of trait level in adaptive testing: Some proposals. Paper presented in the 2007 GMAC Conference on Computerized Adaptive Testing. Minneapolis.

Rastgoo, A., & Namvar Y.(2010). Assessment Approaches in Virtual Learning, *Turkish Online Journal of Distance Education-TOJDE*, 11(1), 42-48.

Sitthisak, O., Gilbert, L., & Davis, H. C. (2007). *Towards a competency model for adaptive assessment to support lifelong learning*. In: TENCompetence Workshop on Service Oriented Approaches and Lifelong Competence Development Infrastructures, 11-12 January 2007, Manchester, UK.

Tian, J., Miao D., Zhu X., & Gong, J.(2007). An Introduction to the Computerized Adaptive Testing. *US-China Education Review*, 4(1), 72-81.

Triantafillou, E., Georgiadou, E., & Economides, A. A. (2007). Applying Adaptive Variables in Computerised Adaptive Testing. *Australasian Journal of Educational Technology, AJET*, 23(3), 350-370.

Weiss, D. J. (2004). Computerized Adaptive Testing for Effective and Efficient Measurement in Counseling and Education, *Measurement and Evaluation in Counseling and Development*, 37(2), 70-84.